Speaker-Attributed STT

Who Spoke the Words?

EARS Fall 2003 Workshop

SASTT - Overview

□ System Description

□ Performance Analysis

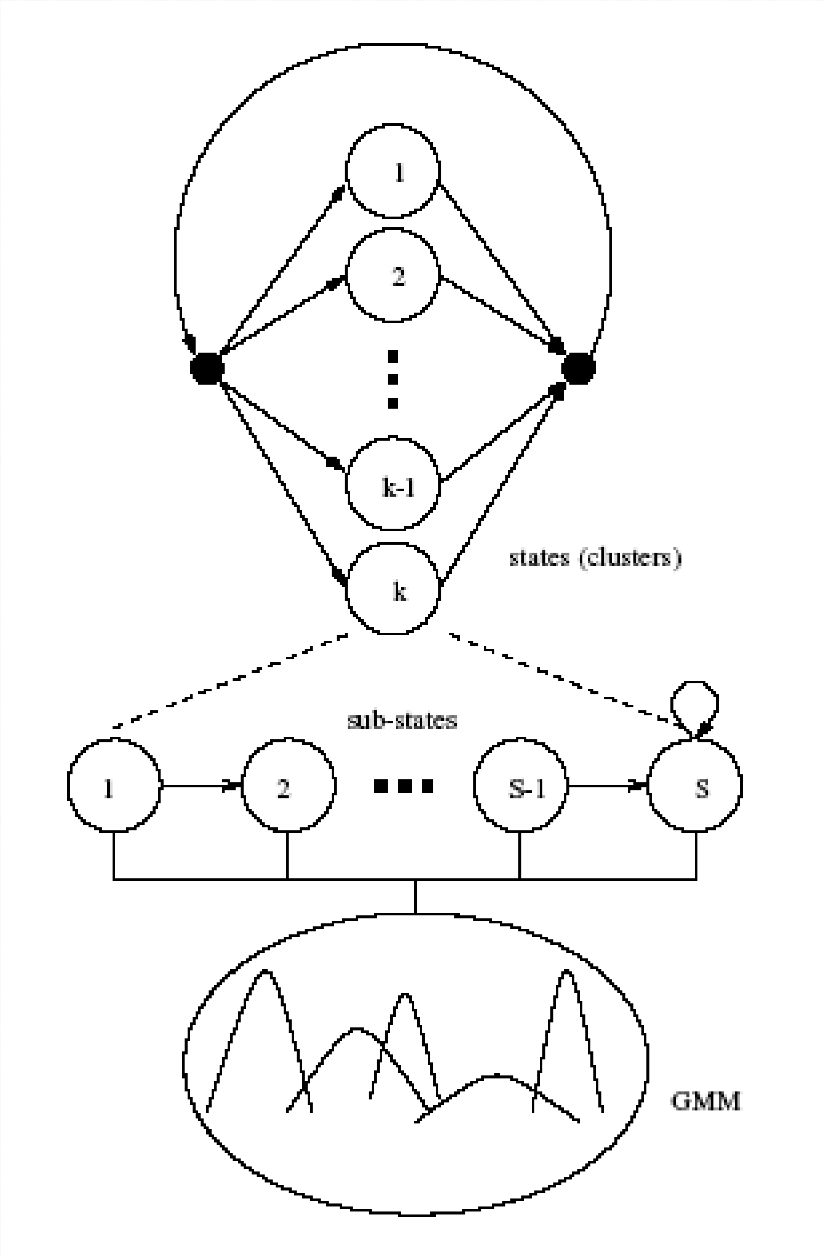
□ Post-Eval Results

□ Two Channel CTS experiments

□ Future work

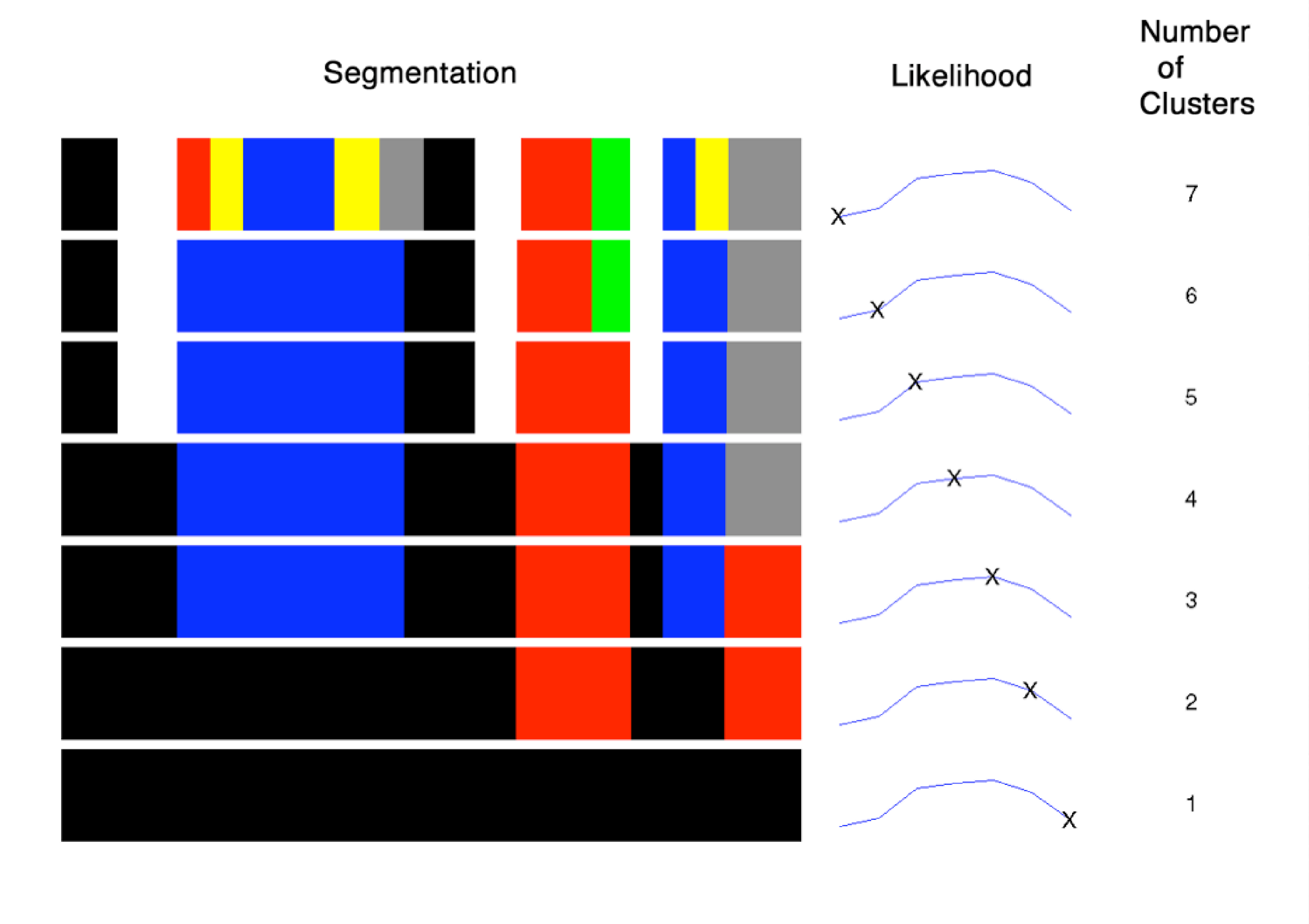
System Description

- □ Basically the same as Spring 03 system
- □ Overlayed words on a "who spoke when" system
- □ Dropped the Speech/ Music detector



System Description (Cont'd)

- □ Agglomerative clustering
- □ Iterative merging of clusters (# params remains fixed)
- □ Stop merging at max in likelihood function
- □ Cluster merging criterion similar to BIC
- □ Details in ASRU03 paper



System Advantages

□ No training data required

□ No "special" tweaking factors or penalty terms

□ Dev data used for system design but...
Need well-matched dev data

Performance Analysis

Observation:

Performance poorer than expected based on Spring evaluation and experiments with dev data.

Official Eval Results			
	SASTT	RT-03	RT1
Ref	15.65	19.07	0.0
Spch	19.46	29.19	12.67

Analysis (Cont'd)

How well would we have done using our Spring 03 system?

		Fall vs. Spring (Eval Data)		
		SASTT	RT-03	RT1
Fall	Ref	15.65	19.07	0.0
	Spch	19.46	29.19	12.67
Spring	Ref	10.35	13.90	0.0
	Spch	14.22	24.25	12.67

What happened?

Analysis (Cont’d)

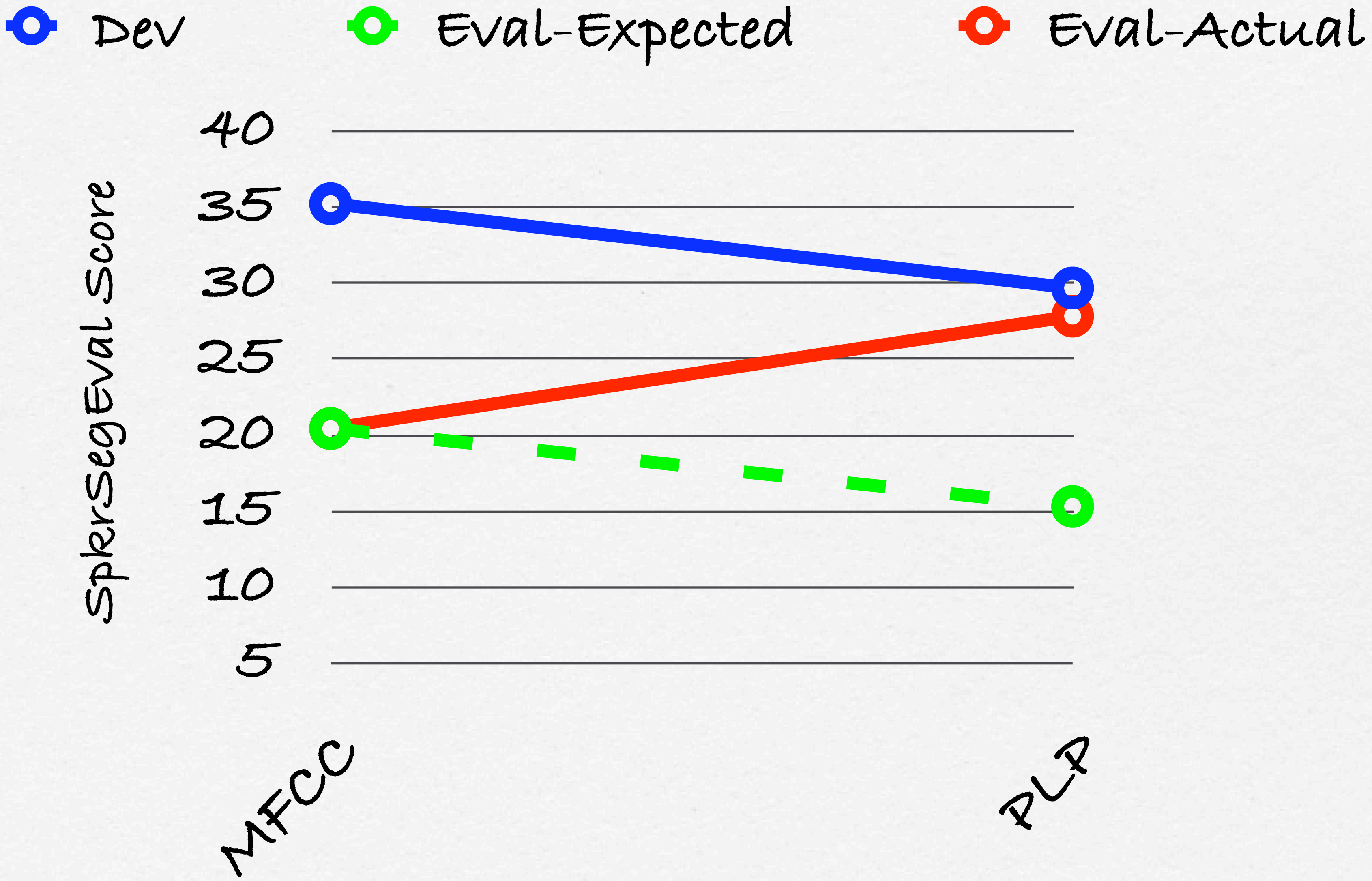
What Changed and Why?

Results on Dev Data		
	SpkrSegEval	rteval (ref)
MFCC19	35.13	24.76
PLP12	29.60	20.17

Switched from MFCC to PLP

Analysis (Cont'd)

What were we expecting?



Analysis (Cont'd)

- □ Dev data poorly matched to the Eval data
  - □ We knew this from our Spring work, but assumed trends would be valid.
- □ Poor performance on Eval data due to the fact the we made decisions about the system based on the Dev data

Post-Eval Results

		SASTT	RT-03	RT1
Ref	Official	15.65	19.07	0.0
	New	9.63	13.21*	
Spch	Official	19.46	29.19	12.67
	New	13.47	23.28*	

*Includes other MDE post-eval improvments

New = MFCC19, and 4sec min dur.

2-Channel CTS Segmentation

- □ Goal: Improve segmentation in cases of cross-talk

  (Is this really a problem? Much of swbd1, 10% of swbd-cell, ? Fisher)

- □ Approach:

  model = parallel GMM's + GLM
  features = single-channel MFCC's, cross-channel corr coeffs & lag, channel-normalized energy ratio

Work by UW

2-Channel CTS Segmentation

□ Preliminary Results Using SRI's 5xRT System

no WER reduction over SRI single-channel algorithm

0.7% absolute gain in oracle single/cross-channel experiment

0.2% absolute gain from preliminary auto-switching algorithm

□ Current Work:

Improve auto-detection of channels with cross-talk

Algorithm refinements and speed-up

Move to HMM framework

Work by UW

Future Work

□ More research on the front-end: believe there is a lot to be gained.

□ For SASTT- Make use of the word timing info for segmentation

The End